Journal of Eye Movement Research
1(1):1, 1-2

Fixation Region Overlap: A quantitative method for the analysis of fixational eye movement patterns

Stephen J. Johnston
Bangor University

E. Charles Leek
Bangor University

This article presents a new method for the quantitative analyses of fixation patterns in eye tracking data. The Fixation Region Overlap Analysis (FROA) uses thresholded spatial distributions of fixation frequency or duration to determine regions-of-interest (ROIs). The locations of these ROIs are contrasted with fixation regions of other empirically-derived, or modeled, data patterns by comparing region pixel overlap. A Monte Carlo procedure is used to assess the statistical significance of fixation region overlap based on 95% confidence intervals (C.I.) of the distribution of random overlap for each set of thresholded ROIs. The value of the FROA method is demonstrated by applying it to data acquired in an object recognition task to determine which of two potential models best account for the observed fixation patterns.

Keywords: Eye tracking, analysis, statistics, methodology

Introduction

The study of eye movement patterns is a powerful tool in vision research. It has provided invaluable insights in a variety of domains including, scene analysis (e.g. Rajashekar et al., 2007; Parkhurst & Niebur, 2003), reading (see Raynor, 1998 for a comprehensive review), visual search and object localisation (e.g. Henderson, 1993). Eye trackers generate several sources of data that permit a variety of analyses of oculomotor responses such as saccades (i.e. rapid, high amplitude, gaze shifts), smooth pursuit (i.e. tracking a moving stimulus) and fixations (where foveal vision is stabilised over a stationary location). Here we focus on the analysis of ocular fixation patterns arising from relatively high amplitude gaze shifts, and consider how quantitative methods may be used to contrast fixation data against both random distributions, and specific model predictions.

It has been estimated that 80% of visual scanning time is accounted for by fixations, (Duchowski, 2003; Manor & Gordon, 2003; van Diepen, De Graef & d'Ydewalle, 1995). Typically, a fixation is defined during preprocessing of gaze data according to specific temporal and spatial thresholds (e.g. Manor & Gordon, 2003). This is necessary because of the continuous micro-saccades, tremor and drift that are characteristic of ocular fixations. These thresholds will specify a minimum dwell time, typically between 100-300 ms, during which successive gaze samples must remain within the area of a predefined image region (e.g. a circle of 1 degree diameter).

One theoretically interesting feature of fixational analyses is their potential to reveal, among other things, properties of psychologically relevant image content at specific display locations. For example, such methods have been widely used to examine eye movement patterns during scene analyses in terms of the visual saliency of particular image locations (e.g. Foulsham & Underwood, 2008; Mannan, Ruddock & Wooding, 1997; Underwood, Foulsham, van Loon, Humphreys & Bloyce, 2006), and in the context of local shape information processing during two-dimensional pattern recognition (e.g. Renninger, Verghese & Coughlan, 2007).

DOI 10.16910/jemr.1.3.5

ISSN 1995-8692

One approach that may be taken is to analyse fixation data (e.g. frequency and duration) in terms of their distributions across regions-of-interest (ROIs) that are defined *a priori*. A fundamental question arising from this approach is how appropriate ROIs are chosen. One simple method is to manually define ROIs based on some hypothesis about theoretically relevant image locations (e.g. particular objects or image features in a scene) and to determine the number of fixations that occur within the defined ROIs. Here, the validity of the resulting predictions is necessarily limited both in terms of the kinds of image features or locations that can be reliability identified by an experimenter, and by the difficulty of incorporating an estimate of subject variation and measurement noise in the ROIs that are defined. This, in turn weakens the reliability of any quantitative comparison between observed and predicted fixation patterns. Alternatively, ROIs may be defined algorithmically in relation to a specific theoretical hypothesis about image content at specific locations. For example, such ROIs could define regions of high contrast, or visual saliency, based on the distributions of low-level image statistics in the displays such as visual saliency based on low-level image statistics (Itti, Koch and Neibur, 1998; Itti & Koch, 2000; Walther & Koch, 2006). This approach has been taken in a variety of eye tracking studies, for example, of visual saliency effects during scene analysis (e.g. Foulsham & Underwood, 2008; Underwood et al, 2006). A further example is provided by a study conducted by Mannan and colleagues (Mannan et al., 1997) who recorded fixation patterns while observers examined scenes in either their original forms, or in low- or high-band pass filtered versions. The ROIs, generated for comparison with the recorded fixation patterns, were selected based on the results of analyses that calculated a variety of image parameters (e.g. luminance maxima and edge density) for the observed scenes. An important advantage of this approach is that it allows for the definition of predicted fixation ROIs based on complex, algorithmically specified, image properties thereby removing the problem of subjective decision making in ROI placement. However, one potential limitation is the requirement to specify a fixed geometry for the ROIs in order to allow for reliable inferential statistical contrasts between observed and predicted patterns. For example, ROIs must be kept to a fixed size so that the probability of a fixation falling within any one ROI is not biased to those with larger areas. Often this is accomplished via selecting a fixed region size about a peak value (e.g. Mannan et al., 1997). The issue with this approach is that the ROIs size and shape may not reflect the characteristics of the spatial distribution of the underlying image parameters, i.e. we cannot expect that all values of a parameter equidistant from the peak value are equal. To change ROI selection in such a way that inclusion of a point in the image was based on the attainment of a critical parameter value would result in unevenly shaped and sized regions that would render Gaussian based statistics invalid.

Privitera and Stark (2000; see also Fujita, Privitera & Stark, 2007) approached the problem from a slightly different perspective. They compared human fixations with those obtained from a variety of image processing algorithms (IPAs), but for both sets of observations, human and artificial, they merged the individual fixations and IPA peak values into clusters that served as regions of interest for the human ('hROI') and artificial ('aROI') data, respectively. The degree of similarity between the observed and algorithmically derived clusters was then determined and used as a dependent measure in further analyses. This similarity measure used a distance metric to determine whether a predicted, artificial, ROI was equivalent to a human, recorded, ROI, such that generalised regions where both human and artificial observer 'fixated' emerged from the data. Of particular note, this similarity measure could not only take into account the spatial similarity of the ROIs, but also the temporal properties allowing a measure of both general similarity as well as sequential similarity. While this approach is elegant in so much as not only spatial, but also temporal properties are examined, the more general method developed by Privitera and Stark for comparing modelled and human data, the 'global similarity' measure, potentially lacked discriminability. The global similarity measure, the measure that most closely resembles the approach used here, does not feature any mechanism to take into account the distribution of fixations around the locus of an ROI. During the calculation of the global similarity measure, all the individual fixation points in a cluster are used to determine the locus of the ROI but, once the locus is determined, all remaining fixations in the cluster are 'removed'. It is these loci, for recorded and model data, that are compared to obtain a value of global similarity. If two models share a similar pattern of predicted regions of interest, i.e. similar loci, but predict a different size of

DOI 10.16910/jemr.1.3.5

ISSN 1995-8692

ROI, this way of calculating similarity would potentially find no difference. That is to say, no explicit account was taken of the size of the ROIs in the statistical process. Therefore a potentially large aROI, obtained via thresholding one IPA parameter map would gain no statistical advantage over a similarly placed, yet smaller aROI, from a different IPA parameter map, if both are assigned as being the 'same' hROI.

A recent study by Foulsham and Underwood (2008) approached the statistical issues raised by ROI selection in a slightly different way by expanding the statistical models used beyond those offered by traditional Gaussian-based approaches. Model derived ROIs, while still artificially created using a fixed radius circle around peak values, were compared against the empirically observed pattern of fixations; statistical validity was assessed by comparing the similarity of the human fixation data to both model predictions as well as to random fixation patterns. Two random models were created. In the first random model fixations could appear anywhere in the image. In the second 'biased' random model the density of fixations from the experimental task was calculated over a 5x5 grid superimposed on the image. The random placement of each fixation was then biased to fall within each of the 25 boxes, with a probability equal to the experimental data although, within each box, the placement was random. This method introduces the use of model derived population distribution estimation, or Monte Carlo, methodology for the assessment of statistical validity, i.e. assessing the distribution of fixations that is expected from a random model to determine the likelihood that the observed distribution occurred by chance. The benefit of this approach is that the generated random distribution can be significantly different from a Gaussian model yet, provided the model is suitably formulated, can be used to determine the statistical significance of the empirically observed fixation data. However, the full power of this approach has yet to be explored since the Foulsham and Underwood (2008) study still uses the selection of fixed shaped ROIs for the models under test rather than allowing quantitative comparison to ROIs derived from empirical data sets with varying sizes and shapes.

To summarise we suggest that the generation of valid *a priori* model-based predictions for the locations of fixation regions (e.g. ROIs) in eye tracking studies raises a number of important methodological issues which include: the method used to generate the predicted region locations (e.g. manually, algorithmically or empirically), and how noise (e.g. arising from between- and within-subject variation as well as error in the precision of the eye movement measurement) is incorporated into those predictions. Whilst *a priori* methods can be used to control for the chance likelihood of a fixation falling in one area vs. another by arbitrarily creating ROIs of a given size, they may fail to accurately resemble the model data in shape and extent, therefore limiting validity. The approach of Foulsham and Underwood (2008), while demonstrating several advantages over other methods, still necessitates the construction of artificially shaped regions to enable statistical comparisons between modeled and empirical fixation patterns to be made.

Here we outline a new method for the analyses of fixation data that addresses the issue of restrictive *a priori* ROI selection, along with the associated problems of ROI definition, and provides a quantitative and statistically valid means of comparing observed fixation patterns to those predicted by different theoretical hypotheses as well as by random distributions. We first describe the general rationale of the methodology which we refer to as Fixation Region Overlap Analysis (FROA). We then illustrate in detail the use of the FROA technique in a study of fixational eye movements during three-dimensional object recognition (Johnston & Leek, 2008; Leek & Johnston, 2008).

An Overview of the FROA Methodology

Rather than attempting to prescribe ROIs *a priori*, FROA contrasts ROIs generated algorithmically from any number of model-based theoretical predictions (aROIs) to empirically defined ROIs from human data (hROIs). The derivation of hROIs is achieved by applying a 2-D Gaussian smoothing function to the filtered collapsed gaze data (e.g. fixation frequency or fixation duration). The smoothing function produces a distribution of fixation frequency or duration data of a specified pixel width (sigma). This smoothed distribution has the advantage of summarising the individual fixations to a sub-sampled

DOI 10.16910/jemr.1.3.5

ISSN 1995-8692

representation where generalised regions become evident and, in which the width of the smoothing function is used to incorporate a noise estimate stemming from within- and between-subject variation (e.g., from saccadic drift) as well as eye tracker resolution. hROIs are obtained by appropriately thresholding the smoothed data such that, for example, the most frequently fixated regions are identified. These hROIs are uniquely suitable for quantitative analyses since their size and shape is directly determined by the fixation data from which they are drawn. Just as with the hROIs, the aROIs can be generated by appropriately thresholding maps of model parameter estimates to create regions of interest. Models of eye movements can then be tested by assessing the degree of spatial overlap between the hROIs, derived from the observed fixation data, and those predicted by aROIs. In FROA, the statistical significance of this overlap is determined by the generation of a constrained 'random' distribution of 'fixation' hROIs with the aROIs to assess the likelihood of the overlap between model data and observed data occurring by chance. The model that describes the random distribution can be constrained to reflect both stimulus-driven and natural biases in scanning such as centre-of-gravity effects or COG (e.g. He & Kowler, 1989), where such biases influence which image regions are more likely to be sampled (see below). The significance of the overlap for any actual fixation region is determined with reference to the 95% confidence interval (C.I.) of the random distribution for that stimulus.

FROA – Implementation

In this section we describe the FROA method in more detail. To do so, we use, for illustrative purposes, an artificial data set before showing an analysis that was performed using FROA methodology on a 'real' data set in the final section.

Data Analysis

The FROA method involves several stages as follows:

1. Pre-processing of raw gaze data using a spatial and temporal filter to define ocular fixations.

Raw data are filtered to extract ocular fixations. We have defined fixations using both spatial and temporal thresholds (Manor & Gordon, 2003) according to which eye movements occurring within an area specified by an ellipse with a diameter of 60 pixels for at least 100 ms were treated as the same fixation. Although we have used distance and latency between recorded eye movements to identify discrete fixation event, alternative methods are available, such as the discontinuous saccadic movement between fixations (Privitera & Stark, 2000). In either case, the application of the analysis method is unaffected beyond the requirement to justify suitable parameters as physiologically appropriate.

2. The generation of global fixation region maps for each stimulus. This involves collapsing the filtered gaze data across observers by applying a 2-D Gaussian smoothing function.

The fixation patterns from the behavioural task are used to generate fixation 'region-maps'. The region-maps are a graphical representation of the frequency distribution of fixation data across the stimulus image. Region-maps are created for each stimulus by collapsing the filtered fixation data (from the previous step) across participants and applying a 2-D Gaussian smoothing function of a specified kernel size and width (sigma). Here we used a sigma of 2.5cm (radius), a value that corresponds to the approximate area of focus for the human fovea at a distance of 60cm. As with all smoothing operations, the aim is to improve signal to noise for the analysis; altering the sigma will have implications for the resultant maps, although the location of the central peak of the ROI should remain consistent there will be changes to the extent of the ROI created. As for the previous steps where values are selected based on assumptions of physical and dynamic characteristics, beyond the need to justify such parameters, one of the strengths of the statistical approach used here is its robustness to changes in parameterisation that alter the characteristics of the ROI. Application of the filter yields a 'global fixation' map showing the distribution of fixations across the image. The first 300ms of eye movements post stimulus onset were removed to eliminate early object localisation fixations associated with COG effects (e.g. Denisova, Singh & Kowler, 2006; He & Kowler, 1989; Whitaker, McGraw, Pacey & Barret, 1996). These effects manifest as a con-

DOI 10.16910/jemr.1.3.5

ISSN 1995-8692

sistent early fixation to the centre of each stimulus. It should be noted that since the data during this stage is collapsed across observers, individual variation in the data is lost and therefore the regions of interest show the most frequently fixated points in the image, but these points are not necessarily fixated by all observers.

3. Thresholding the region maps to produce a binary image encoding the maxima of fixation frequency or fixation duration for each stimulus.

From the global fixation map, a binary thresholded region map is obtained for each stimulus. This procedure has been implemented here using a script written in Matlab (Mathworks inc.) using the image processing toolbox. These binary threshold maps are created by thresholding the global fixation map to define those regions showing the highest fixation frequencies. For consistency across stimuli the threshold is set here at T = N/2, where N is the total number of participants. Thus, for example, in the later exemplar study of object recognition, a region must receive a minimum of 12 fixations (given 24 participants) to be classified as a frequently fixated region (see Figure 5). This value is essentially an arbitrary figure that can be altered, but must obviously remain consistent within a single experiment. The value of N/2 has been used since, anecdotally, it show good levels of intra-stimulus correlation of fixation patterns yet discernable levels of inter-stimulus variation. Again, as with the definition of fixations, the statistical approach is robust to changes in the thresholding parameter – lower thresholds that lead to more regions surviving fixation will require higher levels of fixation overlap in order to achieve significance so the choice of thresholding is free within the bounds of feasibility. An image is then created where all the pixels that exceed the thresholding level are assigned a greyscale value of 255 (white), while all other pixels are assigned the value 0 (black). Once thresholded, the maps are binarised such that the 'most frequently fixated regions' are shown in white while the remainder of the image is black, i.e. binary maps.

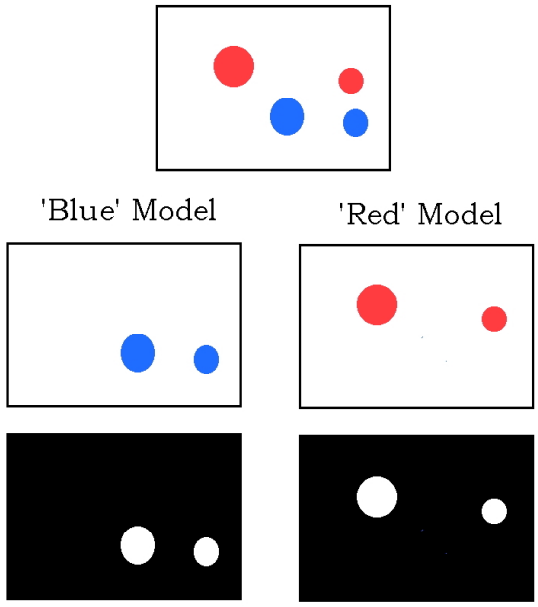

Figure 1 – Exemplar data for two models, 'Red' and 'Blue', showing predicted areas where the models expect the most frequently fixated regions to occur. Top panel shows both models together, middle panel shows the models separately, and the lower panel shows the binarised map that is used as the models input to FROA.

4. The determination of pixel overlap between observed thresholded fixation region maps and the model under test.

The next step is to determine which of the model patterns best accounts for the pattern of fixation regions found in the recognition memory task. This is achieved by converting all of the thresholded region-maps for the model data from each data set into binary (black and white) maps similarly to the participants gaze fixation data. Figure 1 shows a hypothetical example of two models, a 'Blue' and a 'Red' model. Each model predicts a region of the display where the most eye fixations are expected to occur. These models are thresholded and, as with the gaze fixation data, formed into binary maps. These binary maps, along with the binary maps from the gaze fixation data are used in the calculation of pixel overlap. The number of shared (i.e. overlapping) pixels between the binary maps of the model and recorded data sets is then computed. Region pixel overlap is the

DOI 10.16910/jemr.1.3.5

ISSN 1995-8692

principal measure used in FROA to assess the degree to which the models can explain the pattern of fixational eye movements in the original task data. Again, we have implemented this routine in a Matlab script that computes the total number of shared pixels in the thresholded images. This process is shown diagrammatically in Figure 2.

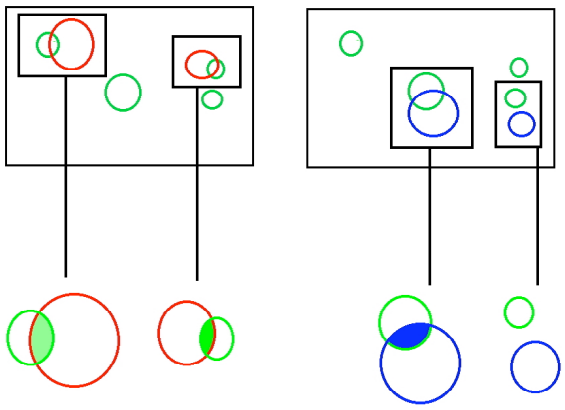

Figure 2 – Figure depicting the 'overlap', the dependent variable used in the analysis. Shaded areas show those regions where there is overlap; the number of pixels in the shaded region is the measure of similarity between model, aROIs, and participant, hROIs.

5. Calculation of statistical significance of observed region overlap relative to the confidence intervals of the distribution of 'random' region overlap with the model.

The validity of any contrast depends upon taking into account variation in thresholded fixation region location, size, shape and orientation for each stimulus between the observed data and modeled data sets. If this were not done a major consequence would be the biasing of the overlap measure towards region maps containing a greater number of pixels. In an extreme case, an image in which the region maps contained 100% of pixels in the stimulus image would always account for 100% of overlap for smaller fixation regions in any contrast data set for the same item. We have already described how, in previous steps, there is a degree of latitude in selecting appropriate parameters because of the way FROA calculates the statistical significance of the overlap. To address this FROA estimates the distribution of overlap expected had the placement of hROIs, obtained from the eye movement recordings been a random process. This procedure utilises the hROIs obtained for each stimulus from the recorded data thereby controlling for region size, shape and orientation. This is done by initially deconstructing the thresholded maps into bounded fixation regions, i.e. separating each discrete cluster of activation and saving it as a single unit. This is again implemented in Matlab; each ROI is extracted on the basis of being a 'closed' bounded region. The centroid of each region is then determined and the ROI saved by calculating the distance of each pixel in the ROI relative to the ROI's centroid. In this way we save the ROIs in a manner that makes it simple to re-insert back in a random location. To replace the ROI, a co-ordinate is randomly generated and the ROI is reconstructed in its new position by placing the centroid at the random position. The random placement of fixation regions is further constrained such that the centroid of each region must fall with the bounding contour of the 'feasible' image regions. The 'feasible' regions include all the points in the image where a fixation may be expected to occur. The determination of the 'constrained' region from the total stimulus display can be obtained in the same way as the ROIs are extracted. Any region that is considered feasible can be added into a mask such that the valid regions for ROI replacement onto the stimulus display are black while invalid regions are white. The pixel locations of the black regions can then be obtained and the placement of the hROIs back onto the image would be restricted to those regions that are within the confines of the masked region of the stimulus display.

A Monte Carlo procedure (Mooney, 1997) is used to generate the distribution of random overlap by taking each hROI and relocating it within the masked regions of the original stimulus, as described above. After each iteration the number of overlapping pixels between the hROIs in the random locations and the model data, the aROIs, is calculated, using the same Matlab procedure that calculates the overlap of the real data, hROIs, with the model data; the random replacement procedure is repeated 1000 times. This generates a constrained distribution of random fixation region overlap for any given data set and stimulus. Figure 3 shows three iterations of the Monte Carlo procedure for example gaze fixation data with the 'Blue' model from Figure 1. In this example we see the four hROIs that survived the thresholding process being shifted about the image and at

DOI 10.16910/jemr.1.3.5

ISSN 1995-8692

each step the overlap with the model data is calculated. The statistical significance of a particular contrast is then determined by comparing the actual pixel overlap in the contrast of interest (e.g. between a given stimulus in the recognition task and the visual saliency map for the same image) and comparing this to the 95% confidence interval of the overlap of the random distribution with the same contrast of interest (e.g. with the visual saliency map).

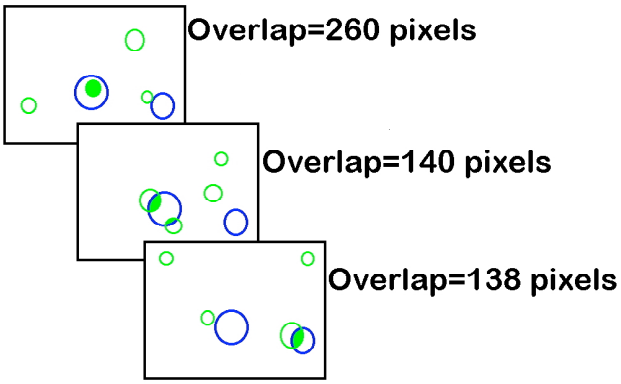

Figure 3 – A diagrammatic example of the calculation of the critical value of overlap, the 95-percentile point of the distribution of overlap values for aROIs with randomly replaced hROIs. Here three iterations of the minimum 1000 iterations are shown. The blue circles correspond to the areas that are predicted to be most fixated by the 'blue model', while the green circles represent the hROIs

FROA – An illustrative example: Analysing fixation patterns during object recognition

Here we illustrate FROA methodology to analyse fixation data collected from a study of object recognition (Leek & Johnston, 2008). We employed two different methods to generate predicted fixation patterns for two potential models of local shape feature analyses during three-dimensional (3-D) object recognition. The first model tested the visual saliency hypothesis (Itti, Koch & Neibur, 1998) using the Saliency Toolbox implementation in Matlab (Walther & Koch, 2006). This essentially produces a weighted contrast map based on low-level image statistics for intensity, colour and orientation. The algorithm was used to generate a saliency map for each stimulus (see below). The saliency maps were thresholded in the same way as the original fixation data from the recognition task. This model represents the fixation patterns we would expect if fixations were the result of eye movements to the most visually salient image regions (see Figure 4 'Model 1'). The second model generated predicted fixation regions based on the locations of 3-D segmentation points at surface intersections producing negative minima of curvature (e.g. Cohen & Singh, 2007; Hoffman & Richards, 1984). For simplicity we refer to this at the '3-D segmentation model'. One way of generating the predicted fixation region locations for this model, as in the case of the visual saliency hypothesis, would be to use a computational implementation that detects negative curvature minima from a 2-D or 3-D model of the stimulus (e.g. Sukumar, Page, Gribok, Koschan & Abidi, 2006). Instead, we generated predicted fixation patterns for the 3-D segmentation model using a trained observer technique in which observers were instructed (after training) to fixate only image regions containing an intersection between two surfaces that form negative minima of curvature (see Figure 4 'Model 2'). This 'trained observer' technique provides a method of generating predicted fixation patterns that necessarily incorporate subject variation and measurement noise that are more comparable to those naturally present in data collected from another experimental task. A limitation is that the technique can only be used to generate predicted patterns for types of image features (e.g., corners, edges, surface intersections, coloured regions) that can be reliably detected by observers. Fixation region maps for both models were generated according to FROA using the same filtering, Gaussian smoothing and thresholding as used for the original fixation data from the recognition task. The Monte Carlo simulation was constrained such that viable regions of the image for random replacement of hROIs were within the bounds of the object undergoing recognition.

DOI 10.16910/jemr.1.3.5

ISSN 1995-8692

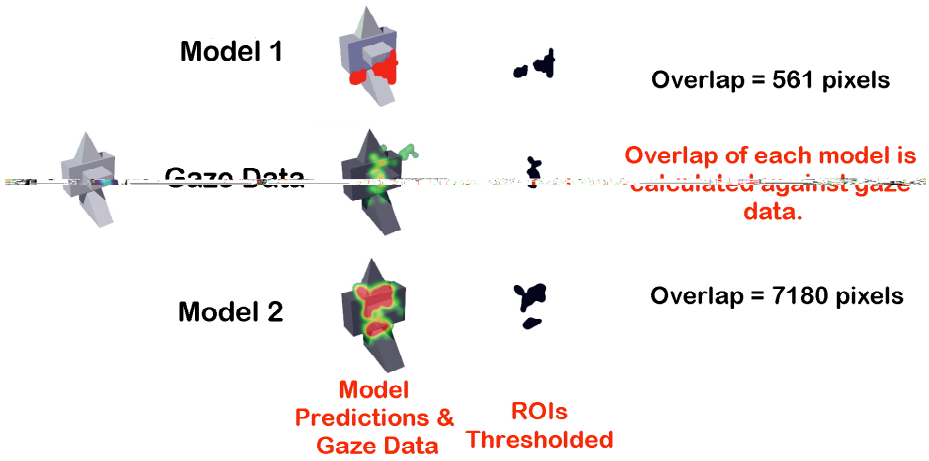

Figure 4 – An example stimulus along with the thresholded-recorded data ('Gaze Data' - middle) and the predicted fixation patterns from the saliency model ('Model 1' - top) and the 3-D intersection model ('Model 2' - bottom). Also shown are the binarised and extracted ROIs along with the overlap of each model's ROIs with the recorded data ROIs

Data Acquisition

Eye movements were recorded on a Tobii ET-17 binocular eye tracker. Data were acquired at a sampling rate of 50Hz with a spatial resolution of 0.5 degrees. Eye position was calculated as the average of the left and right eye positions. Head movement and viewing distance (60 cm) were controlled using a chin rest. Stimuli consisted of 60 (10 objects x 6 viewpoints) surface rendered novel greyscale objects (see Figure 5), illuminated from a single light source in the upper left-hand quadrant, and scaled to fit within a frame of 800 x 800 pixel dimensions equating to 15 degrees of visual angle horizontally. This scale was chosen to induce saccadic exploration around the stimuli. The eye-tracker calibration procedure was as follows. There were 24 participants (Mean age 22.67 years, 22 right handed, 17 female). Participants viewed a static blue dot that appeared, randomly, in each of 16 possible screen locations. From the recorded eye position and known screen position, a transformation matrix was constructed, via a linear interpolation method, which was used to determine gaze position from eye position. The calibration results were visually inspected to ensure a sufficiently good calibration was performed prior to continuing beyond the calibration stage.

Eye movements were recorded while participants memorised (Learning Phase) and then recognized (Test Phase) sets of computer generated 3-D novel objects (see Figure 5). In the learning phase participants viewed five (target) objects each from three different viewpoints (0, 120 and 240 degrees about the image plane). Stimuli were presented at the centre of the monitor sequentially for 10 seconds each following a three second fixation at a peripheral location randomly selected in any of the four corners of the screen. In the test phase, targets and an additional set of five visually similar distracters were presented in a recognition memory task at previously seen (0, 120 and 240 degrees) and novel (60, 180 and 300 degrees) viewpoints. Stimuli were presented centrally following fixation at the corner location. Stimuli were displayed until response. The participants were asked to determine and respond via key-press (k – 'yes'/ d – 'no') whether the presented 3-D object was one of the objects seen in the learning phase of the experiment.

DOI 10.16910/jemr.1.3.5

ISSN 1995-8692

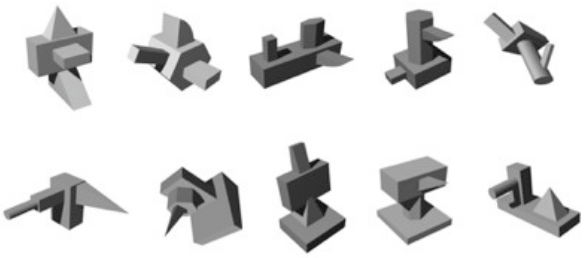

Figure 5 – The ten grey-scale 3-D rendered stimuli that were used in the experiment. For each stimulus shown an additional 5 were created by obtaining images of the same object rotated 60, 120, 180, 240 and 300 degrees about the image plane.

Here we present the results for one of the target stimuli, shown in Figure 4. The overlap for the visual saliency model (total pixels = 23786) with the recognition task data (total pixels = 121086) is 561 pixels (3.7%), while the overlap of the recognition memory task data with the 3-D segmentation model (total pixels = 22057) is 7180 pixels (47.7%). Figure 6 shows the results of the Monte Carlo simulation that we used to assess the significance of the reported overlap. The mean overlap for the 'random' distribution with the visual saliency model was 2456 pixels, against a 95% C.I. value of 7313 pixels. In contrast, the Monte Carlo simulation using the 3-D segmentation model produced a mean overlap of 1853 pixels and 95% C.I. cut-off of 6554 pixels. Given our values for the degree of overlap of the recognition memory task data with each of the models we can conclude that the 3-D segmentation model accounts for a significant amount of fixation region overlap while the visual saliency model does not.

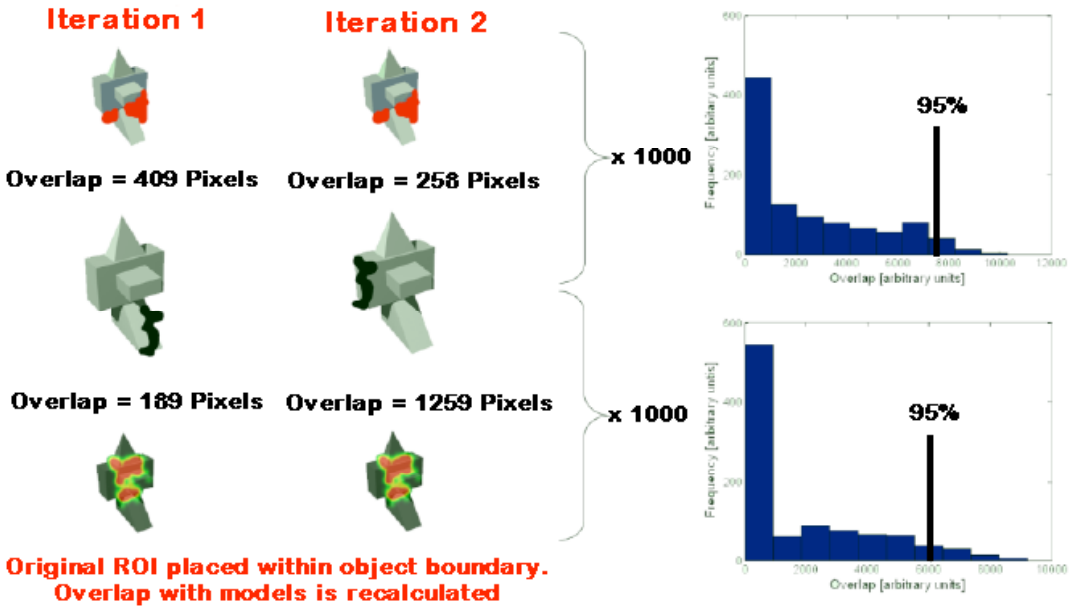

Figure 6 – An example of two of the one thousand iterations performed to build up the distribution of the overlap expected had ROI placement in the recorded data been a random process. The right hand side histograms show the results of the Monte Carlo procedure for the saliency data (top) and the 3-D Segmentation model data (bottom).

DOI 10.16910/jemr.1.3.5

ISSN 1995-8692

In order to determine the statistical significance across stimuli FROA calculates the frequency distribution of significant contrasts relative to the random distribution that would be expected by chance given the number of contrasts performed corrected for multiple comparisons. For example, where there are 60 contrasts (i.e. separate overlap comparisons on 60 stimulus displays) one would expect 3/60 contrasts on a null hypothesis to exceed an alpha level of .05 (that is, N contrasts x α [60 x .05 = 3]). The observed number of significant contrasts can therefore be compared using a χ^2 distribution against the distribution predicted under the null hypothesis given the number of contrasts performed.

For the illustrative data sets, comparisons of fixation frequency region maps showed that 3/60 (10 objects x 6 viewpoints) stimulus contrasts for the recognition task (test phase) versus the visual saliency model exceeded the 95% C.I. of the random distribution, χ^2 (d.f. = 1) = 0.18, p = .67. In contrast, 29/60 stimuli in the recognition task versus 3-D Segmentation model contrast were significant, χ^2 (d.f. = 1) = 28.81, p < .0001. A similar pattern was found for mean fixation duration. Here only 2/60 contrasts in the recognition task versus visual saliency model contrast, χ^2 (d.f. = 1) = 0.00, p = 1.0 and 35/60 in the recognition task versus 3D Segmentation model contrast, χ^2 (d.f. = 1) = 37.01, p < .0001, showed significant overlap relative to the random distribution.

DISCUSSION

In this article we have outlined a new method for the quantitative analysis of the spatial distribution of fixational eye movements. The FROA approach provides a statistically rigorous method for the comparison of empirically generated fixation data with fixation region patterns predicted by different theoretical models. As we have illustrated, the approach can be applied to contrasts between empirically-derived fixation data and theoretical predictions from models using manually defined ROIs, those based on computational simulations of hypotheses or trained observer techniques that incorporate estimates of noise from subject variability and measurement error.

A related approach has recently been described by Foulsham and Underwood (2008). They contrasted predicted fixation locations for visual saliency with observed eye movement patterns for scenes using a recognition memory task involving stimulus encoding (learning) and recognition (test) phases. Fixation patterns obtained during the two experimental phases were statistically compared to generated random models. Two different random models were created to test the significance of the co-occurrence of real and saliency predicted fixations, a 'simple' random model and a biased random model. Foulsham and Underwood's method, although sharing the same underlying principle of testing a predicted distribution with that obtained from a random one, does have one notable difference, the selection of ROIs. In Foulsham and Underwood (2008) the ROIs are created using a fixed radius circle surrounding peak model values whereas we allow our ROIs to be arbitrary in shape and size. Whilst this is appropriate in their case, due to the statistical analysis that was used, it does constrain the models in so far as their ROIs must be equivalent in terms of size at each theoretically interesting point, which may not be the case.

A similar limitation occurs in the approach of Privitera & Stark (2000) & Fujita, Privitera & Stark (2007) who create their model based ROIs such that they become a point in the image space where, in the following similarity measure, they are assigned as equivalent to a defined human ROI, or not. In their global similarity measure, the amount of aROIs accounted for as being equivalent to human, pROIs, are then determined for each algorithmic model and contrasted. While this method is strengthened by not requiring the use of a strict distance measure to define a given human ROI as being equivalent to an algorithmically modeled one, since the clustering method used does not require the use of a fixed boundary, it does suffer from not allowing the creation of ROIs to have any size or shape. The method illustrated here can potentially determine different levels of contribution for two different models aROIs, even if they are coincident since the size and shape of the aROI will potentially alter the level of overlap, the primary dependent variable used in our approach. The strength of the model presented here is that where models account for 'more' of an images visual area we can include that in the modeling process without violating statistical rules.

DOI 10.16910/jemr.1.3.5

ISSN 1995-8692

The approach here advances previous techniques in several ways. First, it allows the use of arbitrarily sized and shaped ROIs since FROA makes no assumptions about the Gaussian distribution of region overlap. Second, the FROA approach does not enforce a size and shape constraint on the ROIs. This permits the use of trained observer methods for the derivation of predicted fixation regions that incorporate noise characteristics, and which are particularly appropriate where computational implementations for the model under examination are not available. It is also worthy of note that both the FROA approach described here, and the methodology described by Foulsham and Underwood (2008), use a constrained random model in order to assess the significance of the observed fixation pattern. Failure to do so will result in a biased estimate of the amount of overlap that would be expected by chance and render the inferences less valid.

One aspect of the approach used within FROA to generate random distributions of fixation regions is that the distributions are highly skewed and non-Gaussian. This provides further support for the use of the Monte Carlo procedure, and assessing statistical significance relative to the 95% C.I, rather than using parametric inferential statistical measures which assume a normal distribution. In addition, as we might intuitively expect, as the number of pixels in the model increases so does the threshold for the 95% C.I: for the visual saliency model, with 23786 pixels the 95% C.I. is 7313, while for the 3-D segmentation model with 22057 pixels, the 95% C.I. is 6554. This illustrates how the FROA approach provides a valid basis for quantitative contrasts between fixation region data sets that vary in region size. The analysis also shows how this method can successfully detect differences between the amount of gaze data that different models can explain beyond simple differences in the number of pixels in the model data. For example, despite the small difference (1729 pixels) between the number of pixels, and therefore the area, of the two hypothetical models used here, the difference in the amount of gaze data that they account for is very large, 561 pixels for the saliency model versus 7180 pixels for the 3-D segmentation model.

We now aim to extend this method in to not only take into account the spatial distribution of fixation data, but also the temporal sequence of those fixations. Fixation sequence information has been investigated previously (Foulsham & Underwood, 2008; Fujita, Privitera & Stark, 2007; Privitera & Stark, 2000; Mannan, Ruddock & Wooding, 1997; Mannan, Ruddock & Wooding, 1995), extending this work to account for arbitrary shaped and sized ROIs would provide a similar level of freedom in selection of the ROIs than is allowed for with current approaches. A second future direction will involve incorporating the pixel-by-pixel viewing frequency data into the model. In its current form, the analysis treats all areas within the regions of interest as being equally weighted in terms of their significance; i.e. the thresholding process removes information regarding which areas of the ROI were fixated more often than others. We aim to extend the method to incorporate measures of the 3-D spatial distributions of fixations across image space. This will provide a means of contrasting hROIs and aROIs at a finer spatial scale taking into account local maxima and minima in the fixation distributions within ROIs.

Additional Information

Example Matlab routines for performing FROA, along with an example data set are available from the author's website http://www.bangor.ac.uk/~pssc04.

References

Cohen, E.H. & Singh, M. (2007). Geometric determinants of shape segmentation: Tests using segment identification. *Vision Research*, 47, 2825-2840.

Denisova, K., Singh. M. & Kowler, E. (2006). The role of part structure in the perceptual localization of a shape. *Perception*, 35, 1073-1087.

Foulsham, T. & Underwood, G. (2008). What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition. *Journal of Vision*, 8, 1-17.

He, P. & Kowler, E. (1989). The role of location probability in the programming of saccades: Implications for 'centre of gravity' tendencies. *Vision Research*, 9, 1165-1181.

DOI 10.16910/jemr.1.3.5

ISSN 1995-8692

Henderson, J.M. (1993). Eye movement control during visual object processing: Effects of initial fixation position and semantic constraint. *Canadian Journal of Expermental Psychology*, 47, 79-98.

Itti, L. & Koch, C. (2000). A saliency-based search methanism for overt and covert shifts of visual attention. *Vision Research*, 40, 1489-1506.

Itti, L., Koch, C. & Neibur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, 1254-1259.

Johnston, S. & Leek, E.C. (2008). Fixation region overlap analysis (FROA): A data driven approach to hypothesis testing using eye gaze fixation data. Poster presented at 8th *Annual Meeting of the Vision Sciences Society*, Naples, Florida.

Leek, E.C. & Johnston, S. (2008) Fixation locations during three-dimensional object recognition are predicted by image segmentation points at concave surface intersections. *Platform Talk at the 8th Annual Meeting of the Vision Sciences Society*, Naples, Florida.

Mannan, S., Ruddock, K.H. & Wooding, D.S. (1995). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 565 – 572.

Mannan, S.K., Ruddock, K.H. & Wooding, D.S. (1997). Fixation patterns made during brief examination of two-dimensional images. *Perception*, 26, 1059-1072.

Manor, B.R. & Gordon, E. (2003). Defining the temporal threshold for ocular fixation in free-viewing visuocognitive tasks. *Journal of Neuroscience Methods*, 128, 85-93.

Mooney, C.Z. (1997). *Monte Carlo Simulation (Sage University series on Quantitative Applications in the Social Sciences, 07-116).* Thousand Oaks, CA: Sage.

Parkhurst, D.J. & Niebur, E. (2003). Scene content selected by active vision. *Spatial Vision*, 16(2), 125 – 154.

Privitera, C.M. & Stark, L.W. (2000). Algorithms for Defining Visual Regions-of-Interest: Comparison with Eye Fixations. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 22(9), 970 – 982.

Rajashekar, U., van der Linde, I., Bovik, A.C. & Cormack, L.K. (2007). Foveated analysis of image features at fixations. *Vision Research*, 47, 3160 – 3172.

Raynor, K. (1998). Eye movements in reading and information processing. *Psychological Bulletin*, 124, 372 - 422

Renninger, L.W., Verghese, P. & Coughlan, J. (2007). Where to look next? Eye movements reduce local uncertainty. *Journal of Vision*, 7(3):6, 1 – 17.

Sukumar, S., Page, D., Gribok, A., Koschan, A. & Abidi, M (2006). Shape measure for identifying perceptually informative parts of 3D objects. *Proceedings of the Third International Symposium on 3D Data Processing, Visualization, and Transmission*, 679-686.

Underwood, G., Foulsham, T., van Loon, E., Humphreys, L. & Bloyce, J. 2006. Eye movements during scene inspection: A test of the visual saliency hypothesis. *European Journal of Cognitive Psychology*, 18(3), 321-242.

Walther, D. & Koch, C. (2006). Modeling attention to salient proto-objects. *Neural Networks*, 19, 1395-1407.

Whitaker, D., McGraw, P.V., Pacey, I. & Barret, B.T. (1996). Centroid analysis predicts visual localization of first- and second-order stimuli. *Vision Research*, 36, 2957-2970.

DOI 10.16910/jemr.1.3.5

ISSN 1995-8692